DOCUMENT RESUME

ED 211 836 CE 031 181

AUTHOR	Payne, Thomas A.
TITLE	Conducting Studies of Transfer of Learning: A Practical Guide. Final Report.
INSTITUTION	Dayton Univ., Ohio. Research Inst.
SPONS AGENCY	Air Force Human Resources Lab., Brooks AFB, Texas.
REPORT NO	AFHRL-TR-81-25
PUB DATE	Jan 82
CONTRACT	F33615-77-C-0054
NOTE	33p.
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Flight Training; Guidelines; Guides; Job Training; *Laboratory Training; Literature Reviews; Military Training; Postsecondary Education; Research Design; *Research Methodology; Research Utilization; *Transfer of Training
IDENTIFIERS	*Air Force

ABSTRACT

This document is a guide for use by the practical researcher concerned with conducting studies of transfer of learning from pretraining of pilots in ground-based environments to performance in aircraft. While the material addresses principally transfer of learning of pilots, many of the issues should be applicable to other contexts, to include training of other aircrew members or even individuals who have quite different tasks to perform. The paper does not deal with theory but, rather, is concerned entirely with method of the transfer study. Method issues, including the planning, task, students, performance measurement, instructors, and analyses, are central to arriving at precise estimates of transfer effects--approaching as closely as possible the maximum that might have been demonstrated, providing a goal for the operational instructor pilot. Study models discussed include those for percent transfer of learning and for the transfer effectiveness ratio. Use of the latter should be essential in providing answers to contemporary questions concerning how much simulator pretraining can be used to replace aircraft training time--without reducing the production of the combat effective pilot. The guide was developed through review of published and unpublished studies of transfer of training from ground-based simulator to actual plane flying done during the past twenty or more years. Eleven steps were identified and are sequenced in the guide for use by researchers. (Author/KC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

AFHRL-TR-81-25

CE

AIR FORCE HUMAN RESOURCES LABORATORY

ED211836

CONDUCTING STUDIES OF TRANSFER OF LEARNING: A PRACTICAL GUIDE

By

Thomas A. Payne
665 Terraine Avenue
Long Beach, California 90814

As Consultant to:

University of Dayton Research Institute
Dayton, Ohio 45469

OPERATIONS TRAINING DIVISION
Williams Air Force Base, Arizona 85224

January 1982

Final Report

Approved for public release; distribution unlimited.

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

CE031181

AIR FORCE SYSTEMS COMMAND
BROOKS AIR FORCE BASE, TEXAS 78235

NOTICE

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely Government-related procurement, the United States Government incurs no responsibility or any obligation whatsoever. The fact that the Government may have formulated or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication, or otherwise in any manner construed, as licensing the holder, or any other person or corporation; or as conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

The Public Affairs Office has reviewed this report, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This report has been reviewed and is approved for publication.

MILTON E. WOOD, Technical Director
Operations Training Division

RONALD W. TERRY, Colonel, USAF
Commander

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFHRL-TR-81-25	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* CONDUCTING STUDIES OF TRANSFER OF LEARNING: A PRACTICAL GUIDE		5. TYPE OF REPORT & PERIOD COVERED Final 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Thomas A. Payne		8. CONTRACT OR GRANT NUMBER(s) F33615-77-C-0054
9. PERFORMING ORGANIZATION NAME AND ADDRESS University of Dayton Research Institute Dayton, Ohio 45469		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 62205F 11230236 11230337
11. CONTROLLING OFFICE NAME AND ADDRESS HQ Air Force Human Resources Laboratory (AFSC) Brooks Air Force Base, Texas 78235		12. REPORT DATE January 1982 13. NUMBER OF PAGES 30
14. MONITORING AGENCY NAME & ADDRESS *(if different from Controlling Office)* Operations Training Division Air Force Human Resources Laboratory Williams Air Force Base, Arizona 85224		15. SECURITY CLASS. *(of this report)* Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT *(of this Report)* Approved for public release; distribution unlimited.		
17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)* flying training percent transfer of learning pilot training training for transfer transfer effectiveness ratio transfer of learning transfer of learning study method transfer of training transfer performance measurement		

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

This is a guide for use by the practical researcher concerned with conducting studies of transfer of learning from pretraining of pilots in ground-based environments to performance in aircraft. While the material addresses principally transfer of learning of pilots, many of the issues should be applicable to other contexts, to include training of other aircrew members or even individuals who have quite different tasks to perform. The paper does not deal with theory but, rather, is concerned entirely with method of the transfer study. Method issues, to include the planning, task, students, performance measurement, instructors, and analyses, are central to arriving at precise estimates of transfer effects—approaching as closely as possible the maximum that might have been demonstrated, providing a goal for the operational instructor pilot. Study models discussed include those for percent transfer of learning and for the transfer effectiveness ratio. Use of the latter should be essential in providing answers to contemporary questions

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

Item 20 (Continued):

concerning how much simulator pretraining can be used to replace aircraft training time—without reducing the goodness of the end product, the combat effective pilot.

SUMMARY

Objective

The objective was a practical guide for use in conducting studies of the transfer of learning from training in a flight simulator to performance in an aircraft.

Background/Rationale

Studies of transfer of learning usually have the goal of providing information about the effectiveness of training techniques and/or equipment for use in designing or upgrading training programs. The likelihood that the information will be used depends on the extent to which both study method and results are convincing in the eyes of the operational user. Studies demonstrating large performance effects resulting from simulator pretraining certainly will be the most convincing and, other things being equal, will be the most likely to promote the adoption and use of new training techniques or equipment during operational flight training.

During the past three decades, numerous studies have investigated the effects of training in ground-based flight training devices on subsequent performance in the aircraft. These studies have employed a variety of experimental techniques. Some of the techniques used were scientifically sound, while others were methodologically flawed and resulted in findings of questionable validity. This diversity of approaches probably resulted in large part from differences in the scientific sophistication or applied research experience of the investigators, as well as conditions peculiar to the specific settings in which the studies were performed. A review and consolidation of the lessons learned from previous studies should be beneficial in guiding future efforts towards increased validity and practical utility.

Approach

The approach used was to review published and unpublished information on transfer of learning and experimental design relevant to pilot training. This information was then carefully analyzed to identify the key issues and factors that must be considered in order to conduct useful transfer-of-learning studies in a flight training environment. Finally, a sequence of steps to be followed by the practical researcher in conducting credible studies was developed and put in guidebook form.

Specifics

The concept of transfer of learning is defined in the guide as any measurable effect of training in a prior task on performance in a subsequent task. The procedures of the typical transfer study are described, and two measures of transfer of learning (i.e., percent transfer and the transfer effectiveness ratio) are defined. Initial discussion of the transfer-of-learning study emphasizes the importance of planning. The remainder of the report identifies and describes 11 steps to take in performing a successful transfer-of-learning study.

The first step is definition of the immediate problem. Its importance is illustrated by asking and considering the answers to a number of questions that serve to focus and sharpen the definition of the research problem. Selection of the task or tasks to be trained is the second step identified. Criteria for selecting the training tasks are suggested. In addition, reasons for identifying research resource requirements early in the study are pointed out.

The third and fourth steps involve the determination of what learners should be involved in the study and the identification of appropriate performance measures. A number of critical aspects of these issues are discussed, including the composition of the sample of learners, their assignment to study groups, and the development of objective performance criteria to serve as a basis for evaluating the learner's performance in the simulator and in the aircraft.

The use of the instructor as a research participant, and, how to plan sufficient time for the study, are the fifth and sixth steps. The seventh step involves the avoidance within a study of factors that may dilute transfer of learning. Advanced scheduling and the need for planning the study to be run in the midst of normal flying training operations are emplasized in steps eight and nine.

Step ten, testing the methodology before collecting final data, and step eleven, the analysis of the data, conclude the presentation of the procedures for conducting a transfer-of-learning study.

Conclusions/Recommendations

This guide provides the practical researcher with valuable guidelines for conducting studies of transfer of learning from training in a simulator to performance in aircraft. In addition, the guide is applicable to a variety of synthetic pretraining environments, including a mix of ground training facilities such as audio-visual media, part-task trainers, and relatively sophisticated simulators.

It is recommended that the guide be given wide distribution in both the training research and operational training communities.

PREFACE

This report was prepared under Consulting Agreement RI-81923 (Revised) with the University of Dayton Research Institute with Dr. Harold D. Warner as project director. This report is a segment of a larger University of Dayton Research Institute effort conducted under contract F33615-77-C-0054 with the Operations Training Division of the Air Force Human Resources Laboratory, Williams AFB, Arizona. The report represents a portion of the on-going work within the Air Combat Training Research Subthrust, and specifically the Flying Training Specialized Support and Data Base Integration component. The associated Project Vanguard planning summary mission area is Support and Technical Base development.

TABLE OF CONTENTS

		Page
I.	Introduction and Purpose	5
II.	Models of the Transfer of Learning Study	5
	Percent Transfer of Learning	5
	The Transfer Effectiveness Ratio	7
III.	The Importance of Planning	7
	Preliminary Work ("Testing")	7
	Designing for Maximum Possible Estimates of Transfer	8
IV.	The First Step: Definition of the Immediate Problem	8
V.	The Second Step: Definition of the Task	9
	Transfer of Learning for What Phase of the Curriculum?	9
	What Specific Tasks will be Involved?	9
VI.	Assessment of Resources: An Iterative Process	10
VII.	The Third Step: Which Students will be Involved in the Study?	10
	Size of the Student Sample: Representative of What Population?	10
VIII.	The Fourth Step: What Performance Measurement Technique?	12
	Relationship with Tasks: Validity of Performance Measurement	12
	The Sequence: Tasks/Criteria/Limits Allowable/Performance Measurement	12
	Illustration: Number of Trials (and/or Errors) to Performance Criterion	12
	Illustration: Performance Grading	15
	Some Questions	16
	Relatively Molar Performance Measurements	17
	Recording Techniques	17
	Automated Performance Measurement Systems	17
	Performance Measurement in the Aircraft and in the Simulator	18
IX.	The Fifth Step: The Instructors	18
	The Instructor as the Researcher	19
	Training for Transfer	19
	Use of Relatively Simple Aids	20
	Rigorous Adherence to the Study Design	20
	No Instruction During Measurement of Student Performance	21
	Balance of Instructors Between or Among Control and Experimental Groups	21
	The Same Instructor: Simulator and Aircraft	21
X.	The Sixth Step: Planning for Sufficient Study Time	21
	Estimates of Necessary Performance Time	22

Table of Contents (*Continued*)

	Page
XI. The Seventh Step: Avoidance of Dilutant Factors	22
Avoid Time Delays Between Simulator Pretraining and Retraining in Aircraft	22
Pretrain Using the Simulator in Meaningful Blocks of Tasks	22
Colocation of the Simulator at the Site of Airborne Training	23
XII. The Eighth Step: Importance of Scheduling in Advance	23
XIII. The Ninth Step: Plan for Running the Study in the Midst of a Busy Operational Training Environment	24
Cooperation and Support of the Unit Commander and the Unit Operations Officer	24
Planning for Minimum Interference with the Operational Schedule	24
XIV. The Tenth Step: Testing the Study Method Before Taking Final Data	24
XV. The Eleventh Step: Analysis of Results	25
XVI. Some Closing Remarks	26
References	27

CONDUCTING STUDIES OF TRANSFER OF LEARNING: A PRACTICAL GUIDE

I. INTRODUCTION AND PURPOSE

This report has been prepared for use by the practical researcher who is concerned with studies of transfer of learning from pretraining of pilots in a simulator to their performance in aircraft. The expressions "transfer of learning" and "transfer of training" tend to be used nearly interchangeably. Although the distinction may be somewhat trivial, the former is used here since it is the learning, not the process of training, that may transfer from work on a prior task to performance on a second. Also, while the term "simulator" is used here for purposes of brevity, it is not intended to be restrictive in nature but, rather, can be considered to refer to various types of synthetic pretraining environments—frequently a mix of classroom facilities, audio-visual facilities, part-task trainers, and relatively sophisticated simulators. While much of the language in this report will refer to pilots, flight simulators, and aircraft, many of the issues should be applicable to other contexts, including training of other aircrew members or, for that matter, training of individuals who have quite different tasks to perform.

The report will not deal with theory (such as the question of what transfers) because such issues are covered elsewhere. The concern will be entirely with method of the transfer study, including the consequences of failure to follow empirically derived principles. The material stems principally from the experiences of the author and his associates, beginning with their work under guidance of the late Professor Alexander C. Williams, Jr. who directed pioneer studies at his original Aviation Psychology Laboratory of the University of Illinois. The report submits techniques and lessons learned from experience, dating perhaps from 1949 when few prior rules were available to the researcher. Descriptions of many of the techniques were not included in early papers for various reasons, and still other techniques may have been considered too obvious to note. Over intervening years, however, it has become clear that many of the issues are not at all obvious, and since they nave been of great service in a number of previous studies, the intent here is to make them available to others concerned with transfer research.

Issues of research method to be discussed have been found essential during attempts to arrive at estimates of transfer that are precise—approaching as closely as possible the maximum that might have been demonstrated during a particular study. Studies of transfer of learning are fragile in the sense that a study that ignores too many issues of method is likely to lead to inconclusive results. Such inconclusive results are serious because they can lead to disinterest on the part of both the research community and the operational training community—disinterest in factors such as new instructional techniques or special aspects of equipment used in the study. The resulting disservice is clear, considering that a carefully planned and conducted study might have led to entirely different types of results supporting concepts that might have been used with considerable value to the research and training communities.

At first glance, the transfer-of-learning study can appear deceptively simple when actually it is not. The number of important issues can be legion, and the precision of subsequent results depends on the compounding effects of many factors.

II. MODELS OF THE TRANSFER OF LEARNING STUDY

Percent Transfer of Learning

"Transfer of learning" is defined here as any effect of learning resulting from pretraining on a prior task (or set of tasks) upon performance in a subsequent task (or set of tasks). Such a transfer effect, if it exists at all, could be facilitating in nature—comparative performance data suggesting positive transfer—

or it could be interfering in nature—comparative performance data suggesting negative transfer. Let us assume at the outset that the carefully planned and conducted study will be concerned with a positive transfer effect.

While various formulas have been offered for use in the percent transfer of learning model (Ellis, 1965; Gagne, Forester, and Crowley, 1948; Murdock, 1957) only one will be considered here. The model makes use of a control group of students (who are not pretrained on a prior task and whose performance data on a subsequent task serve as a standard) and one or more experimental groups of students (who are pretrained on a prior task and whose performance data on the subsequent task are compared to those of control students for purposes of estimating any transfer effect realized). For the purpose of this study, the prior task(s) may be carried out in a simulator (or other synthetic training environment), with the subsequent task(s) being carried out in an aircraft. The model is

$$\frac{\bar{C} - \bar{X}}{C} (100) = \text{percent transfer of learning}$$

where:

$\bar{C}$: an average of trials, time or errors accumulated by a control group of students to arrive at a performance criterion in the aircraft.

$\bar{X}$: an average of trials, time, or errors accumulated by an experimental group of students to arrive at that same performance criterion in the aircraft, having been pretrained to a performance criterion in a simulator.

Thus, using illustrative numbers:

$\frac{10 - 5}{10} (100) =$ .50-percent transfer of learning. If those values represent hours of training in an aircraft, pretraining of experimental students in the simulator resulted in a 50-percent saving in aircraft training time—on the average.

The numerator of the percent transfer of learning formula would have to be reversed if measurement were in terms of performance grades, such that higher values represented better performance, thus:

$$\frac{\bar{X} - \bar{C}}{C} (100) = \text{percent transfer of learning}$$

where:

$\bar{X}$: an average of grades assigned to experimental students for performance in the aircraft.

$\bar{C}$: an average of grades assigned to control students for performance in the aircraft.

Thus, if students were graded using a 12-point scale (with 12 being superior performance and 0 being total failure), using illustrative numbers:

$\frac{10.50 - 8.75}{8.75} (100) =$ 20-percent transfer of learning. In this case, pretraining of experimental students in the simulator resulted in a 20-percent higher grade than attained by the control students—on the average.

The Transfer Effectiveness Ratio

Recent concern of the pilot training community with increasing costs and shortage of energy led Roscoe (1971, 1972) to state quite a different model. Being concerned with the value of time, the model provides an estimate of transfer effectiveness, using as a standard a measure of the amount of simulator pretraining required by an experimental group of students to evidence superior performance in the aircraft as compared to performance of a control group of students. The estimate can be given by:

$$\frac{\bar{C} - \bar{X}}{S_x} = \text{the transfer effectiveness ratio}$$

where:

$\bar{C}$: an average of trials or time required by a control group of students to arrive at a performance criterion in the aircraft.

$\bar{X}$: an average of trials or time required by an experimental group of students to arrive at that same performance criterion in the aircraft, having been pretrained to a performance criterion in a simulator.

$\bar{S}_x$: an average of trials or time required by the experimental group of students to arrive at a performance criterion in the simulator. Thus, using illustrative numbers:

$\frac{10 - 5}{5} = 1.0$ the transfer effectiveness ratio. If those values represent hours of pretraining in the simulator and hours of training in the aircraft, respectively, 1 hour of pretraining in the simulator saved 1 hour of retraining in the aircraft—on the average.

As can be seen, the difference between the estimate of the percent transfer of learning and the transfer effectiveness ratio is that the former ignores the amount of pretraining required in the simulator, and the latter takes that factor into account. Contemporary questions concerning how much aircraft time might be replaced with simulator time could be addressed principally through studies using the transfer effectiveness ratio.

A later section of this report will consider the problem of the time required for the transfer study, noting that the transfer effectiveness ratio model may suffer more from insufficient time to complete the study. Further, since data necessary for the transfer effectiveness ratio model can be used to compute percent transfer of learning estimates, there may be occasions when it would be of value to use both of these models in the same study.

III. THE IMPORTANCE OF PLANNING

It seems likely that more studies of transfer of learning do not succeed because of inadequate planning and preliminary work than because of any other factor. The study must be planned carefully if results are to be of any real and practical value, and both planning and the study take time. During the planning phase, a sound investment in time is necessary to carry out the work to be described here and to identify and correct or adapt to the problems and the less than optimal limiting factors that may be imposed by real-world constraints.

Preliminary Work ("Testing")

As is the case with any formal study that costs time and money, the study of transfer of learning should not be conducted without sound preliminary information that suggests the type of outcome likely

to be found. The formal study should not be conducted in an exploratory manner to establish trends or directions of findings, but rather, it should be conducted to arrive at an estimate of the magnitude of a transfer effect. It should be concerned with reasonably substantial effects that could be of practical significance in the real world—not with statistically significant trivia.

Trends, directions of findings, or the likely existence of a positive transfer effect should be established during one or more relatively simple tests from which ideas, hunches, or hypotheses evolve. While the precise nature of such preliminary work will depend on the particular problem of the moment, in some cases early testing might be fairly simple, using only a few students, relatively simple equipment, and perhaps relatively crude performance measurements. Preliminary "mini-studies"—assuming that they involve reasonable care—can be invaluable, particularly if several experimental students who have been pretrained in some specific manner seem to show dramatically superior performance in the air as compared to performance of several control counterparts. Information obtained in this way can lead to a highly useful formal study.

Among the other valuable insights that might be provided by preliminary testing, deficiencies of the simulation equipment could result in negative transfer effects. Preliminary work can help to identify such problems, together with a means for solving them; in this case, planning for the process of training for transfer—a subject to be discussed in a subsequent section of this report.

Designing for Maximum Possible Estimates of Transfer

The goal of the researcher should be to plan and conduct a carefully controlled study, taking every possible precaution in the design to ensure that the resulting estimates of transfer are precise—that is, that they approach as closely as possible the maximum levels that could be demonstrated. Because of uncontrollable variables, research-demonstrated techniques could result in less than optimal transfer effects when used in an operational training program, still the researcher should attempt to demonstrate the maximum possible transfer effects to show what can be accomplished and thereby provide a goal for the operational instructor. Without knowing what could be done, the operational instructor could tend to be satisfied with lesser results.

IV. THE FIRST STEP: DEFINITION OF THE IMMEDIATE PROBLEM

Although the underlying question concerns the extent to which prelearning in a simulator will transfer to performance in an aircraft, the first step should involve consideration of the specific purpose of the particular transfer study. Various specific purposes can have different associated problems such as the following.

Will the study be concerned with combat readiness of experienced pilots facing reductions in aircraft time for skills maintenance and reacquisition training? Prior to asking whether lost aircraft time might be replaced with simulator training, preliminary work should have to do with an assessment of degrees of combat readiness. Is there evidence of decay of skills with reduction of aircraft time?

Will the study be concerned with effectiveness of basic pilot training in the face of reductions in aircraft time? Prior to asking whether aircraft time can be replaced with pretraining in a simulator, it would be well to be sure that effectiveness is actually reduced.

Will the study be concerned with experienced pilots in transition to a new type of aircraft and mission? A preliminary question should ask whether there exist facilities that are truly adequate for pretraining work.

Will the study be concerned with pilots returning to flight duties from predominantly administrative assignments? Again, are there facilities that are truly adequate for pretraining work?

Although much of contemporary interest in using simulator pretraining is motivated by concerns with costs of aircraft time and the energy problem, the nature o the synthetic training environment is such that it can provide benefits over and beyond those of saving money or fuel. Does the purpose of the study involve one or more of the following issues?

A well designed simulation facility can be used on an all-weather, 24-hour basis and as a substitute when training aircraft are not available. In addition, it can provide a safe training environment; it can be used to compress time during training, enabling concentration upon critical segments of flight tasks rather than requiring that time be lost while flying to and from a practice area; and it can provide opportunities for observation and measurement of student performance that ordinarily are not possible in the air. The student can be interrogated easily on the spot concerning reasons for errors, and exercises can be rendered standardized and repeatable, affording very precise assessments of learning progress. In the event that the specific purpose of the study involves one or more of these issues, perhaps the major concern lies with the measurement of percentage of transfer of learning rather than with arriving at an estimate of transfer effectiveness.

In any event, it seems important that the researchers have identified all aspects of the purpose of the transfer study being conducted.

V. THE SECOND STEP: DEFINITION OF THE TASK

Transfer of Learning for What Phase of the Curriculum?

It is impracticable to attempt to measure transfer of learning for an entire curriculum through a single study. Thus the study is likely to be concerned with a specified phase of a training curriculum, such as training for takeoff, approach and landing, instrument flight, attack on a ground target, air-to-air attack using a weapon-control subsystem, or other meaningful phase that has continuity. In some cases, it might be that even a particular phase is too complex to be dealt with in its entirety, requiring study of one or more segments. If it is desired to arrive at transfer estimates for several phases of a curriculum, it may be necessary to establish their order of priority.

Decisions in this context must depend on requirements of operational organizations, and necessary background details must originate from those organizations. The contributions of highly experienced instructor pilots are very important during the early planning stage, and some studies may require contributions on the part of additional operationally experienced pilots who are not necessarily instructors.

What Specific Tasks will be Involved?

At the outset, the research team must derive definitions of tasks the student will be expected to perform in the operational situation represented in the study. Precisely how this is to be done will depend on the nature of the particular study. Past work has made use of operational sequence diagrams and pictorial diagrams of flight tasks. If the curriculum phase has been selected with care, use of such analytical techniques should result in a convenient number of tasks that can be defined fairly tightly.

The instructor pilot can be of great help during this work by noting high frequency errors that have been made in the past, task segments that are of time-critical nature, and cues that appear to be necessary and sufficient in facilitating performance. These concepts will be considered further during discussion of performance measurement techniques because it is essential that measurement and tasks be related closely.

VI. ASSESSMENT OF RESOURCES: AN ITERATIVE PROCESS

After arriving at a reasonably thorough set of task definitions, the research team must be certain that resources available will enable conduct of the study. That question has to be addressed continuously as planning progresses. Will available simulators be adequate for use during pretraining for the specified tasks? Will pertinent aircraft—in which "proof of the pudding" performance measurements must be taken—be available and in sufficient numbers? Will an instructor cadre be available and in sufficient numbers? Will students of the necessary type be available in sufficient numbers? Will it be possible to run a carefully controlled study in the midst of a busy operational training schedule? Will there be problems in getting necessary support from the commander and the operations officer of the training organization? Will all of these enabling factors continue to be available during the time required to carry the study to completion?

Insufficiency of too many enabling factors could render conduct of the study infeasible or at least could impose serious constraints on what can be accomplished. Thus the research team would do well to keep in mind the question of adequacy of available resources during the entire planning process.

VII. THE THIRD STEP: WHICH STUDENTS WILL BE INVOLVED IN THE STUDY?

It may be that the question of which students will be involved in the study can be answered by the nature of the immediate problem and the nature of the curriculum phase and tasks of interest to the study. Earlier, four categories of pilots were mentioned: pilots requiring skills maintenance and reacquisition training for combat readiness, students in basic flight training, experienced pilots in transition to a new type of aircraft and mission, and pilots returning to flight duties from predominantly administrative assignments. Clearly those categories of pilots represent at least four very different populations—probably far more than that.

Sometimes the researcher may be tempted to extrapolate the transfer study data as far as possible, perhaps wanting to arrive at more information than actually is feasible. The notion of mixing students representative of several different populations of pilots in a single study is a case in point. But if that is done, with the total sample of students being only of modest size, it is unlikely that results could be applied to specific training situations. The rule should be to keep the student sample as homogeneous as possible—particularly when only small samples are available.

Size of the Student Sample: Representative of What Population?

The most frequently asked question may be that of sample size but, unfortunately, there rarely seems to be a truly satisfactory answer. Perhaps the most useful approach is to try to keep the sample(s) as representative as possible of a population of interest.

Ideally, the control and experimental students should be matched in terms of experience and aptitude for the tasks at hand, but in reality, the notion of what "experience" really means is imperfect, and the training research community would appear to have few truly useful tests of aptitude for specific tasks likely to be involved in transfer studies. The total number of flight hours logged probably plays a role in a definition of "experience," but there is at least some empirical evidence that this is by no means an entirely useful predictor of performance levels.

It seems popular to state that the sample size should be as large as the situation permits and, in one sense, that is probably correct. If, in an extreme case, every member of a particular pilot population could be sampled, the accuracy of the predictions concerning transfer would be vastly improved. But that is sheer fantasy, and in the practical world researchers usually have to make do with relatively small samples, the sizes of which are limited by time, funds, and student availability. However, there is no magic in large samples. A small sample composed of highly representative students is likely to yield information of considerable value, whereas a large sample that is either heterogeneous in nature or is

characterized by a bias of some kind is likely to yield misinformation. Further, any effect—as a transfer estimate—that requires very large samples to show itself is unlikely to be of practical significance (Hays, 1973 pp 419—429; McNemar, 1940).

Efforts to match student samples between or among control and experimental groups in past studies have had to be made using a great deal of common sense and in terms of types of students available. Some researchers have used a combination of length of experience and experience in specific types of aircraft, attempting to place equal numbers of such students in the several groups.

If the total available supply of students appears to be reasonably homogeneous—if at least there is no specific reason to predict an imbalance of aptitudes and skills—perhaps the best that can be done is to assign students to the several groups on a purely random basis. The principal concern, of course, is that, if predominantly more apt students are assigned to a control group, a spuriously low transfer effect is likely to be demonstrated, and conversely, if predominantly more apt students are assigned to an experimental group, the demonstrated transfer effect is likely to be exaggerated. So, if relatively small groups must be used—perhaps 8 to 12 students per group—how severe is the problem?

Suppose that a total of 16 students were available, 8 being assigned to an experimental group, such assignments being made at random because there was no real reason to suspect serious differences in aptitude. Suppose further that the 16 students actually were ordered in aptitude for the task at hand but that there was no way to estimate that ordering. This means that eight of the students are the more apt, and with luck, four of them would be assigned to each group. A problem would arise if all eight or seven or six or five of the more apt students had been assigned to the same group. So, binomial probability can be used to estimate the chances of that happening.

$$P(r/..., p) = \binom{n}{r} (p)^r (q)^{n-r}$$

where by definition, $p = q = .5$.

a. The probability that all eight of the major apt students had been assigned to the same group is about .004 (4 chances in 1,000).

b. The probability that seven of the more apt students had been assigned to the same group is about .03 (3 chances in 100).

c. The probability that six of the more apt students had been assigned to the same group is about .11 (11 chances in 100).

d. The probability that five of the more apt students had been assigned to the same group is about .22 (22 chances in 100).

e. The sum of these probabilities—the probability that eight or seven or six or five of the more apt students had been assigned to the same group—is about .36 (36 chances in 100).

While it is realized that this illustration involves a somewhat simplified set of assumptions (it does not, for example, take into account the relative aptitude ranking of the eight more apt students), it does serve to suggest that the probability of absolutely mismatched groups is quite low ($p = .004$) and that the range of probabilities—from seriously mismatched to moderately mismatched groups—is about .03 to .22. These are fairly good odds in favor of a reasonably well matched group. What is more, if the study actually does involve a sizable transfer effect, that effect should show itself even under the less favorable of these situations.

VIII. THE FOURTH STEP: WHAT PERFORMANCE MEASUREMENT TECHNIQUE?

Relationship with Tasks: Validity of Performance Measurement

While earlier work concerned with definition of the tasks will have placed reasonable bounds on the transfer study, more detailed definitions of the student's tasks have to overlap work for development of the performance measurement technique. While the absolute nature of the performance measurement technique will depend on many aspects of the particular study, it is essential that tasks and measurement be related logically. To the extent that such a relationship is established well, validity of performance measurement will just about take care of itself.

The Sequence: Tasks/Criteria/Limits Allowable/Performance Measurement

Although means for expediting the process are likely to differ from study to study, it seems reasonable that consideration of the sequence to be illustrated may be central to establishment of a necessary bridge between task definition and measurement. The sequence implies the following steps:

Define the Tasks Operationally—Exactly what will the student be required to do? Depending on the complexity of the tasks, this may be defined at various levels of detail.

Set Criteria for Performing the Tasks—How are these criteria established by physical facts of the tasks?

Specify Deviations from Those Criteria That Can Be Tolerated—In the same kinds of terms used to define the tasks and performance criteria, what performance *limits* likely will permit of successful completion of the tasks?

Structure the Performance Measurement Units and Means for Taking Data—It is at this point that the process is likely to become iterative, the question being whether desired types of data can be taken.

Illustration: Number of Trials (and/or Errors) to Performance Criterion

The sequence can be illustrated with an example from an early study concerned with transfer of learning in the context of making approaches and landings (Payne, Dougherty, Hasle, Skeen, Brown, and Williams, 1954). Experimental students were pretrained for the task in a simulator, where they were required to achieve a performance criterion prior to moving to the aircraft. Their performances and those of their control student counterparts were measured during retraining in the aircraft. The study used a measurement of the number of trials and errors accumulated before arriving at a total task performance criterion. The illustration to follow is concerned only with performance in the aircraft (the sequence used with experimental students in the simulator having been nearly identical but somewhat attenuated because of limitations of that device).

Definition of the Task (abbreviated here)—The instructor positioned the aircraft for a 90-degree side approach from the left, giving control to the student at this point. The student was required to make necessary power reductions, the turn onto the final approach, the approach proper, the flare, and the touchdown for a wheel landing. The task ended after the aircraft executed a short posttouchdown roll.

(For convenience, performance criteria, performance limits, and the performance measurement process are illustrated in tabular form.)

Performance Criteria	Performance Limits	Performance Measurement
From starting position to position on windline (within imaginary extensions of runway edges):		
1. Airspeed: 90 mph	+10 to -5 mph.	observed on instructor's airspeed indicator.
2. Turn onto approach was not overshot:	Did not pass windline; turn completed within runway width (150 ft).	Observed by instructor from rear seat.
3. Turn onto approach was not undershot:	Did not fail to reach windline; turn completed within runway width (150 ft).	Observed by instructor from rear seat.
4. Aircraft was on windline prior to passing airport boundary fence.	Was within windline (150 ft).	Observed by instructor from rear seat.
5. Student was assisted in no way.	None.	Instructor did not assist student verbally or by control action.
From position on windline to position over end of runway:		
6. Airspeed: 90 mph	+10 to -5 mph.	Observed on instructor's airspeed indicator.
7. No S-turns outside of windline.	Did not depart from windline (150 ft).	Observed by instructor from rear seat.
8. Manifold pressure at 15 in. Hg.	+ 5 in. Hg.	Observed on instructor's manifold pressure indicator.
9. Glidepath aimed at a definite point within first third of runway.	A point between near end of runway and the one-third marker.	Observed by instructor from rear seat.
10. Aircraft crossed near end of runway at 100 ft altitude.	+50 ft.	Observed on instructor's altimeter.
11. Student was assisted in no way.	None.	Instructor did not assist student verbally or by control action.

Point of touchdown:

12. Touchdown executed withia first third of runway.	A point between near end of runway and the one-third marker.	Observed by instructor from rear seat.
13. Touchdown executed in center of runway.	At least one wheel within two white center lines.	Observed by instructor from rear seat.
14. Student was assisted in no way.	Aircraft touched down on main wheels; student allowed aircraft to roll (or to skip lightly) to demonstrate that no serious bounce would take place.	Instructor did not assist student verbally or by control action.

Several points are of interest:

a. The 14 sets of criteria, limits, and measurements were developed during a great deal of preparatory work. The task was carried out using the AT-6 aircraft, with power settings and airspeed being standard for the type of approach and landing used (then called a "transport landing"). (Simulator pretraining work with experimental students used the 1-CA-2/AT-6 Link Trainer, modified to provide a dynamic projection of the runway image.) Task limits were established by the instructors while observing from both the aircraft and the ground. The glidepath angle was measured using a surveyor's instrument—a theodolite—enabling establishment of points for beginning the maneuver and flying the approach with 90 (statute) mph, 2,000 rpm, 30 in Hg of manifold pressure, gear and full flaps down. Thus the subtasks and their performance limits were judged to be entirely valid descriptors of successful execution of the maneuver.

b. The instructor said nothing during each of the student's trials. As the student performed a trial, the instructor made necessary observations and entries for the 14 performance units. Only after repositioning the aircraft for starting a subsequent trial did the instructor make comments and corrective remarks. Had instruction taken place during a trial, the measurements could have reflected those remarks as well as the student's performance—the two being confounded absolutely.

c. A successful approach and landing were defined as the student's having met all 14 subcriteria; missing even a single item was defined as an unsuccessful trial. In this study, the instructor scored performance as it occurred, the process having been possible because of the tandem, two-place aircraft used. Observations were recorded using a standard, knee-clipboard form.

d. The student met total task criterion performance at the point of having made three consecutive successful approaches and landings. Preliminary work had indicated that such performance was highly unlikely on the basis of chance alone. (Tests had shown that once this "three-in-a-row" criterion was met, the student tended to execute a long series of successful maneuvers before a subsequent "out-of-limits" observation occurred.)

e. Some of the subcriteria for successful performance in terms of individual units were of relatively subjective nature and, sometimes, were difficult to score. (Windline examples are a case in point.) It was found necessary to impose a rule that the instructor give a "within-limits" score for any measurement unit about which there was any doubt. Preliminary work indicated that, using this rule, observer-observer

reliability of scoring approached unity. While the rule had the effect of widening acceptable performance limits somewhat, the total measurement technique proved to be highly sensitive to differences in goodness of performance.

f. This technique provided the study with several different types of estimates of transfer of learning. While the principal interest involved percent transfer in terms of number of trials to performance criterion and number of errors during trials to performance criterion, it was also possible to estimate first-trial transfer in terms of errors and to estimate transfer in terms of errors made during the first five trials.

g. In 1953, when the study was conducted, primary interest was with the percent transfer of learning—not the transfer effectiveness ratio. Unfortunately, records that would have enabled calculation of a transfer effectiveness ratio (long after the fact) were lost. Those records showed number of trials to performance criterion for the experimental students during simulator pretraining.

Illustration: Performance Grading

The process of establishing tasks, criteria, limits, and performance measurement can be illustrated with an example from a more recent study concerned with transfer of learning in the context of air combat maneuvering (Northrop Corporation, 1976). The study was concerned with the percent transfer of learning for experimental students who had been pretrained in a special simulator, using an instructor grading system because the portion of the training syllabus that could be involved was too short to permit measurement of number of trials to a criterion. The sequence to be described is concerned only with performance in the aircraft.

Definition of Tasks, Criteria, and Performance Limits—Tasks consisted of eight basic maneuvers used in an air combat maneuvering training syllabus. Instructors provided descriptions of these maneuvers, each of which was divided into logical segments, together with criteria and criterion limits for successful performance. Measurement units were based on these descriptions, together with the types of high frequency student errors that had been observed during operational training.

Performance Measurement (Grading)—It was not feasible to grade performance while airborne because of very short durations of critical maneuver segments, together with the high g forces involved. Therefore grading was done on the ground immediately following the training flight. Instructors used standardized grade sheets, showing the several measurement units, and indicated the type of maneuvers used in each engagement. The two instructors who had worked with the student were required to grade measurement unit on a consensus basis.

The Grading Scale—Instructors graded each measurement unit using letter grades of the following scale:

Grades	Definitions	Numerical Equivalents (enabling analyses)
A+		12
A	Superior	11
A-		10
B+		9
B	Above Average	8
B-		7
C+		6
C	Average	5
C-		4
D+		3
D	Below Average	2
D-		1
F	Failing	0

Instructors used the scale in two stages (not being concerned with numerical equivalents). First they rated each unit across the five-point scale: A through F. Second, when they had entered one of the top four categories, they were asked to qualify the grade as necessary to express their judgment with greater precision (as: B+, B, or B—). This resulted in a highly sensitive 12-point scale that permitted the fine differences in performances to be discriminated.

This type of grading scale has been used in a number of different study contexts, in each case proving highly successful for quantifying expert professional judgment. In this particular study, it was necessary to observe two precautions. First, since there was a marked difference between capabilities of the student pilots and their highly skilled instructors, those instructors regarded the entire range of the grading scale as representing types of student performance only. Second, the five basic grading categories were defined (e.g., the "superior" category represented performances of the top 10 percent of students of the operational training program). Use of such types of definitions seems advisable in an attempt to standardize interpretations of scale categories.

Some Questions

During the process of defining tasks, performance criteria, allowable limits, and measurement units, it might prove useful to ask questions such as the following:

a. Can the tasks be categorized according to segments that have logical start and end points? Do the tasks involve equipment limitations (stall speed, g limits)?

b. At each readily defined, critical mission segment, what is the *crux* of successful performance? Is the judgmental factor or the motor factor the more critical, or are they of equal importance?

c. How is time critical and at what points? Since it is neither possible nor desirable to attempt to measure every aspect of performance, is it possible to associate performance measurement units with time-critical periods or segments of the maneuver or mission? These periods are, after all, when serious errors are most likely to take place.

d. Is it possible to measure problem detection latency? Can it be inferred on the basis of subsequent action?

e. Is it feasible to match "time available" with performance time? Time available for an action, a maneuver, or a mission segment would have to be derived from operational definitions. This kind of performance measurement would appear to be particularly pertinent in terms of combat mission segments. Did the student do the correct thing but take too long to do it?

f. At each readily defined, critical mission segment, is it possible to list the types of errors that students frequently have tended to make in the past?

g. Is it possible to delineate a reasonably small number of aircraft actions or positions involved in carrying out tasks—these being placed in descending order of desirability? Particularly in cases involving single-place aircraft, this may prove to be an essential measurement category—the instructor having to make a judgment from a position in another aircraft.

h. Is it possible to estimate the student's level of concentration? This might involve the use of secondary tasks in an attempt to estimate the amount of effort required by the student. Aspects of tasks permitting, the student approaching a high level of learning should have more time and energy remaining for executing additional tasks.

Relatively Molar Performance Measurements

It is suggested that the researcher should not necessarily avoid measurement of performance in relatively molar terms as long as the measurement units are anchored to clear definitions of important tasks, clear definitions of what the student will be required to do, and clear definitions of consequences of serious deviations from the limits provided. Transfer studies should look for large performance differences that could be of practical significance—not small differences no matter the level of statistical significance. Measurements should not deal with molecular trivia simply because they are easy to define and measure.

Recording Techniques

Past work has made use of both hard copy—forms with pencil entries—and tape recordings. In the main, however, hard copy has seemed to be the more useful. For one thing, the printed scoring or grading form provides a checklist of items to be covered. For another, transcribing or listening to tape contents is severely time consuming. And, depending on the type of recorder used, maneuver g forces can slow down the mechanisms, rendering subsequent playback less than truly clear. Whether technological advances and budgets will permit use of forms or truly useful automatic airborne recording techniques remains to be seen.

Automated Performance Measurement Systems

There would appear to be an unfortunate belief in some quarters that an automated performance measurement system, as such, implies associated validity of data. That is, of course, just not so. Validity of measurement data depends on the anchor to reality and has nothing to do with how the measurements are implemented. It might be useful, however, to consider three services that an automated measurement system might provide—those services possibly solving some problems facing the human data taker.

Reliability—An automated system, being subject to less variability in operation than is the human observer, should provide measurement data of greater reliability in the sense of measuring the same type of event from trial to trial and from student to student. Designing manual measurement techniques having high observer-observer reliability can be difficult.

Span of Surveillance—An automated measurement system could take into account all items it is designed to cover—consistently, not being subject either to distraction or to a limited field of view as is the human data taker. Human data takers obtain most their information visually, with the requirement to timeshare—they simply cannot look in several directions at once. And even though human observers may be required to attend only to a very narrow or highly specified aspect of a visual situation, there is the problem of vigilance error, that is, an observer may look at the correct location but too soon or too late.

Access to Information—Many difficulties in measuring performance are a function of not being able to position the human observer to permit a view of the desired events. Consider the single-place fighter aircraft or even a two-place aircraft in which an observer in a second seat cannot see either the student's control actions or the outside world from the student's point of vantage. For an observer located in a second aircraft, the principal source of information is the dynamic physical positioning of the student's aircraft. That is fine from the standpoint that physical positioning is the end product of the student's decision making and action processes, but it tells the observer little about why errors took place. Those reasons must be inferred. The observer has to make do with the things that can be seen.

To the extent that an automated measurement system could be provided with necessary sensing devices and be mechanized economically and in necessary lightweight and compact form, it might be located within the student's aircraft, solving many of these kinds of problems.

Performance Measurement in the Aircraft and in the Simulator

Most of the discussion thus far has been concerned with measuring student performance in the aircraft. Airborne performance measurements are essential to the study of transfer of learning and provide much of the "payoff" information. But performance measurement during simulator pretraining is important too. During the illustration of models of transfer studies, it was noted that simulator pretraining should continue to a performance criterion. If that is not done, the notion that learning has taken place can be something of an act of faith. Study results will have more meaning if evidence is provided indicating that learning did take place during simulator pretraining. This concept holds for either model for the transfer study, but it may be even more critical for the model concerned with a transfer effectiveness ratio.

IX. THE FIFTH STEP: THE INSTRUCTORS

It has been noted earlier that the role of the instructor pilot is critical to the conduct of the study of transfer of learning. Too frequently in the past this factor has been been recognized fully, insufficient emphasis having been placed on the various important contributions of the instructor. This may have been the case because of undue attention paid to the nature of the simulator; this having tended to overshadow more critical issues. Most researchers tend to be enchanted with elegant equipment, this possibly leading to two dangerous semantic traps.

First, it is customary to speak as though simulators "train"; however, they do not, they never have, and they never will. It is the instructor who does the training. The goodness of design of the simulator may be important in providing the instructor with the necessary training environment, but it seems unlikely that engineering and cost restrictions will allow a type of simulator to be designed that will provide a "work sample" so complete that maximum transfer can occur without superior instruction.

Second, a nearly universally expression is that someone, "received training." That unfortunate phrase suggests that the training process is passive and is something like slicing cheese. (How many slices are necessary?) But anyone who knows anything about the training environment that gets things done knows that learning is an active process. Students cannot sit there "receiving training"; they must take an active role, interacting with both the environment and the instructor.

Perhaps some day, there may be a training simulator environment that uses some modified form of the concept of computer-aided, programmed instruction—no human instructor being involved except for purposes of handing special student problems. But even in such a situation, the instruction will remain the key element. Programmed instruction provided with such an advanced simulator should be based on skills, knowledge, and techniques of a large number of instructor pilots, the basic situation being similar to contemporary effective training but taking advantage of such combined information.

The Instructor as the Researcher

The instructor cadre must participate in the design of the study from the outset, providing information that helps anchor the study to reality, particularly with respect to the nature of the task and the performance measurement technique. But over and beyond that work, the instructor ordinarily will conduct the study in addition to the role of guiding the student's learning. During critical airborne work, the instructor is also the researcher and data taker as well as the safety pilot. What is more, the instructor is the most logical individual to handle simulator pretraining of experimental students.

Training for Transfer

The technique of training for transfer has been shown to be critical when features of the simulation environment may be markedly different from those to be encountered in the air. The simulation environment, by definition, is at variance with the prototypical environment. Because of physical and engineering limitations, sometimes aspects of the synthetic environment may be diametrically opposed to those of the operational situation. In such cases, there can exist a "built-in" effect that likely leads to negative transfer—simulator pretraining possibly providing an interfering effect upon subsequent performance in the air. Further, in some cases it may not be possible to carry out particular sub-tasks in the simulator, even though those sub-tasks are very important in the air.

The process of training for transfer involves identifying and being certain that the student understands the limitations of the simulator as compared to an aircraft, and the instructor is uniquely qualified for this responsibility. It may be necessary to perform a particular function one way in the simulator and another way in the aircraft—as is appropriate to each. The student must know about these differences and why they exist. It has been found useful to explain such differences to the student at frequent intervals—at least prior to and during simulator work and prior to and during airborne work. The more severe the differences, the more frequently they should be pointed out.

To illustrate the concept, early transfer studies used a simulator requiring considerable rudder pedal travel with minimal stick movement to perform a coordinated turn (1-CA-2/AT-6 Link Trainer), while the counterpart aircraft (AT-6) required exactly the reverse—little rudder pedal travel with considerable stick movement (Payne et al., 1954; Williams & Flexman, 1949). While this is a dramatic example of built-in potential for negative transfer, work in those studies showed that, if the problem is made quite clear to the student prior to and during simulator work and prior to and during airborne work, such training *for* transfer completely offsets the potential, the student having little difficulty in either the simulator or the aircraft.

The recent study cited, concerned with transfer of learning in the context of air combat maneuvering, involved no fewer than 20 aspects of the simulation environment that differed importantly from their airborne counterparts (Northrop, 1976). The instructor pilots identified those aspects and had them printed on a sheet in descending order of importance, distributing that sheet to all experimental students. In addition, they emphasized the problems during briefing and debriefing sessions for work in both the simulator and the aircraft (F-4J). The following are some of those aspects:

a. Target detail definition decreases greatly beyond 1 mile, but the target remains as a "light source" out to infinity.

b. Simulator provides more instantaneous g than does the F-4J—at all airspeeds.

c. Simulator departs at 30 to 33 units and usually cannot be recovered.

d. Pulling simulator nose up at high airspeeds is more difficult than in the F-4J.

e. It is very easy to exceed 6g in the simulator.

f. The simulator has large amounts of roll divergence.

g. Buffet effects are less intense in the simulator than in the F-4J.

h. Simulator rudder is too sensitive at slow speeds.

i. Flying ACM in the simulator provides a twilight effect: Is similar to flying at dusk.

Subsequent conduct of the study indicated that the experimental students were well aware of the differences and that they had little difficulty making appropriate adjustments and responses during work in the aircraft. Since the set of differences could have provided a marked built-in potential for negative transfer, it is likely that the ultimate information obtained from the study would have been much less important except for this process of training for transfer.

Sensitizing the Student to Necessary and Sufficient Cues—The process of training for transfer can be of value when cues of different types are available in the simulator and in the air. Although the problem may be less severe with today's higher quality of simulation environments, there may be occasions in which cues found most effective in the operational environment cannot be produced in the simulator. Under such conditions, the instructor would do well to point out differences, noting both those cues that are likely most useful in the air and those that can be used for the same purpose in the simulator. This procedure need not be paradoxical because, frequently, different pilots make use of different sets of cues as aids during performance of the same maneuver; these perhaps depending on their individual preferences. Even the same pilot may use different sets of cues at different times, such as while flying types of aircraft that permit of peculiar angles and extents of view. The pilot makes do with alternatives that serve the same purpose.

Use of Relatively Simple Aids

To aid the instructor during the briefing and debriefing sessions, usually it is a good idea to provide models, photographs, chalkboards, or other items of relatively simple equipment that can be used to illustrate points clearly. Air combat instructor pilots have made heavy use of a pair of simple wooden triangular blocks mounted on the ends of dowel sticks. Use of such rudimentary equipment might sound inelegant, but often it appears to serve the purpose extremely well.

Rigorous Adherence to the Study Design

The transfer study, as any other formal study, must be conducted under highly controlled conditions so that resulting data are not confounded with extraneous events. The goal should be that the transfer study reflect only the results of pretraining in the simulator. To provide for such control, students must work with a common syllabus of tasks carried out in a prescribed sequence, in the absence of free-floating variables such as giving a particular student a special exercise (even though, in an operational situation, that might be the logical thing to do). Such deviation from a prescribed sequence of events could render the resulting data uninterpretable. If the instructors are co-designers of the study, they will be unlikely to deviate from standardized procedures, even inadvertently.

No Instruction During Measurement of Student Performance

The study design should provide that no instruction take place while the student is performing and performance data are being taken. If an instructor makes a comment (even casually) during a measurement trial, the resulting data are likely to reflect that input in addition to (and confounded with) the student's ability level.

Balance of Instructors Between or Among Control and Experimental Groups

One of the surest ways to arrive at biased transfer estimates is to allow imbalance of instructional techniques and styles among groups. The problem can be avoided by providing that each instructor work with equal numbers of students in each group of the study. If this is done, the variable of individual differences among instructors will be balanced and as long as the instructors follow basic agreed upon practices, they are free to explain issues and train according to their own personal techniques that they have developed and found effective for their own particular style.

The Same Instructor: Simulator and Aircraft

It is important that the same instructor train the experimental student in both the simulator and the aircraft. This practice is likely to facilitate the effort to arrive at maximum transfer effects. The instructor who has done the simulator pretraining will have the best possible understanding of the individual student's strong and weak points, being able to estimate what that student did and did not learn during pretraining, and being able to use that knowledge to the best advantage during retraining in the aircraft. Immediately prior to an exercise in the aircraft, the instructor can review important issues with the student, refreshing the student's memory of particular performances in the simulator and mentioning significant differences that exist between the simulated and airborne environments.

X. THE SIXTH STEP: PLANNING FOR SUFFICIENT STUDY TIME

It is very easy to overlook the issue of planning for a study syllabus of sufficient duration that all students will have a reasonable amount of time in which to arrive at an end performance criterion (experimental students in the simulator and all students in the aircraft). Failure to provide sufficient time can result in data of the study being attenuated—not all students' performances figuring into analyses. In the worst case, *no* students would arrive at performance criterion—the study being a total failure or else transfer estimates being dependent on a grading process. The point is, of course, that individual students simply are likely to learn at different rates, requiring different amounts of time to arrive at performance criterion.

The cited study concerned with approaches and landings (Payne et al., 1954) ran into a problem as students were in the final phase of making landings in the aircraft. Students, drawn from an Air Force Reserve Officers Training Corps (ROTC) program, were nearing landing performance criterion when their semester ended, and they had to go away. Only 8 of the 12 students met the landing criterion. Fortunately, four of these were in the control group and four were in the experimental group, permitting a reasonable and balanced estimate of transfer.

The cited study concerned with air combat maneuvering (Northrop, 1976) had to be conducted using an operational training syllabus of such short duration that the use of a trials-to-criterion measure was not possible. In that case, the problem was recognized before the fact, with performance measurement consisting of instructors' grades in lieu of trials-to-criterion. While that permitted reasonable estimates of percent transfer of learning, it was not possible to arrive at estimates of a transfer effectiveness ratio. A

form of transfer effectiveness estimate might have been feasible had the syllabus been of sufficient length that an instructor could have shortened (or omitted altogether) portions of a student's mission segments when, in the instructor's judgment, goodness of performance warranted such action. Even that, however, was not possible. Instructor pilots had pointed out, before the fact, that the syllabus was too short to permit a sufficiently high level of learning to justify any omission of syllabus items. And since that syllabus was set by operational training rules, it could not be adjusted.

Estimates of Necessary Performance Time

In designing the study, the goal would be to provide sufficient time for the least apt student (in either or any group) to complete the work and to arrive at an end performance criterion. Preliminary testing would appear to be the best means of estimating necessary time because tasks, their degrees of relative difficulty, associated performance criteria, and types of students can be quite different from study to study. Even use of preliminary testing might not provide a complete answer, considering that only small numbers of students are likely to be involved. But since the consequences of too little available time can be serious, resulting estimates might have to be padded. It is far better to allow too much time than too little.

XI. THE SEVENTH STEP: AVOIDANCE OF DILUTANT FACTORS

"Dilutant factors" are defined here as practices that can prevent demonstration of maximum possible transfer effects of a study. The concern here is with two dilutant factors that are not necessarily mutually exclusive.

Avoid Time Delays Between Simulator Pretraining and Retraining in Aircraft

While the severity of the problem of time delays between the simulator pretraining and the retraining in the aircraft may be dependent on the nature of the specific study, the issue would appear to be highly critical for tasks that are "volatile" in nature—tasks involving skills highly subject to decay in the absence of practice. This may be illustrated in terms of the study concerned with air combat maneuvering (Northrop, 1976). In that study, unavoidable scheduling restrictions required that experimental students be pretrained in the simulator on a massed basis during a 5 day period, moving to work in the aircraft only after completion of that block of simulator work. For a number of reasons, including the facts that the simulator was located more than 100 miles from the airbase, the press of work of the operational training schedule at that airbase, student loadings, shortages of instructors, mechanical difficulties with aircraft, weather, and interruptions of training schedules because of priorities, delays between simulator pretraining and retraining in aircraft were as long as 4 weeks. The principal priority causing interruption of the schedule involved availability of aircraft carriers for qualification training. Carriers became available only infrequently and had to be used immediately. Observation of goodness of performance in the simulator and resulting transfer effect estimates suggested rather strongly that there was a clear and strong dilutant effect.

Instructor pilots who conducted the study noted that skills of air combat maneuvering are quite volatile in the sense that periods of inactivity of as much as 10 days resulted in noticeable decrements in their own performances. It takes little imagination to estimate the performance decrement for student pilots who had completed the simulated equivalent of only six flights in this context.

Pretrain Using the Simulator in Meaningful Blocks of Tasks

Precisely what a "meaningful block of tasks" might be would depend on the context of the particular transfer study. But again, the issue may be illustrated best in terms of the air combat maneuvering study (Northrop, 1976). Experimental students were pretrained in the simulator for the first 6 flights of a 17-

flight air combat syllabus used in the operational training environment—only those first 6 flights figuring into the transfer study. The flights were designed to acquaint the student with tasks of air combat maneuvering in a sequential order, beginning with basics and progressing to engagement exercises of increasingly difficult nature. The initial flight was for familiarization and involved only a single aircraft. Subsequent flights introduced eight basic maneuvers of the total syllabus, with the difficulty of combat engagements being increased. The instructor, flying the "adversary aircraft" during two-aircraft exercises, began by presenting a relatively easy "mark," but increased the complexity of the performance to the point that, by the sixth flight, the student was "fighting" a relatively skilled "opponent."

Once the simulated equivalents of those six flights had been completed, the experimental students moved to the airbase and began the normal training syllabus as used in the operational squadron. It can only be surmised that pretraining in this blocked manner may have been less than optimally effective in terms of transfer of learning. It seems highly likely that had the experimental students been pretrained for each individual flight and retrained in the aircraft for that flight, the resulting transfer subsequently estimates might have been considerably greater.

It can be reported only on the basis of personal observation that resulting transfer estimates seemed far lower than might have been expected without the compounding effects of these two dilutant factors: (a) delay between simulator pretraining and aircraft retraining and (b) massed training of the sort described. In any event, the lesson seems clear. If a transfer study makes use of clearly functional blocks of simulator pretraining, moving experimental students to the aircraft as soon as possible, the resulting transfer effects should be augmented.

Colocation of the Simulator at the Site of Airborne Training

Probably the best way to prevent delays between simulator pretraining and aircraft retraining would be to locate the simulator at the airbase to be used in the study. Even if this is possible, however, proper scheduling would still be critical. But in the event that the simulator must be located elsewhere, every attempt should be made to transport experimental students to the airbase after they have completed logical blocks of simulator pretraining—getting them into the air at the earliest feasible times. The problem and the solution are easy to state. Expediting the solution must depend on aspects of the particular study.

XII. THE EIGHTH STEP: IMPORTANCE OF SCHEDULING IN ADVANCE

The issue cannot be emphasized too heavily. During early phases of planning, the research team should begin to assess potential scheduling problems and should consider these on an iterative basis as final plans take shape. Even prior to testing the study method, a detailed schedule should be prepared, taking into account times for involvement of students, instructors, simulators, and aircraft. This must not be left to chance.

Cooperation of the unit commander and the unit operations officer will be critical to development and enforcement of the schedule, and here as before, the instructors working in the study should be able to help achieve such cooperation.

Means must be found for preventing visitors from interfering with scheduled study work. Experience has shown clearly that this can be a serious problem. Perhaps it can be solved best through orders issued by pertinent unit commanders. The problem tends to be most severe during simulator pretraining. Simulators—particularly those of elegant nature—tend to attract visitors frequently. If the environment permits, it may be possible to provide for a spectator vantage point that does not interfere with training work. Above all, neither the student nor the instructor should be aware of the presence of visitors, especially when those visitors are of high rank.

XIII. THE NINTH STEP: PLAN FOR RUNNING THE STUDY IN THE MIDST OF A BUSY OPERATIONAL TRAINING ENVIRONMENT

If the study is to be conducted in the midst of a busy operational training environment the cooperation and support of the unit commander and of the unit operations officer are required on the one hand, and planning for minimum interference with the operational training program is required on the other hand. While it would be highly desirable to be able to run these kinds of studies using a dedicated facility, it seems more likely that they will have to make use of operational facilities.

Cooperation and Support of the Unit Commander and the Unit Operations Officer

It is easy for the researcher to lose sight of the fact that the operational people have their own problems, and at best, cooperation with a study effort could be simply an additional annoyance. It may be that major objections can be avoided by making the unit commander and the operations officer parties to the purpose of and planning for the study from the outset. While it might be tempting for the researcher to rely on orders from higher authority—these directing the unit commander to support the research work—it takes little imagination to see that this can be a serious mistake. The research team would be wise to work with the operational people from the very beginning, persuading them of the importance of the study and getting their professional inputs for planning the effort. The instructors can play essential roles here, having close professional ties with the operational unit people. In many cases, preparatory work here can make or break the study.

Planning for Minimum Interference with the Operational Schedule

The research team, working with the operational people, should develop a clear set of plans for preventing all but absolutely necessary interference with the operational work. The interference may consist principally of time required for simulator pretraining of experimental students, but the nature of the study may impose still other requirements, to include modified routines during airborne work, use of research instructors, balancing instructors' work with experimental and control students, and special student briefings and debriefings. But if proper rapport, cooperation, and support have been established at the outset, it should be possible to solve various problems to everyone's satisfaction. There is no way to overemphasize the importance of these issues. The process of solving potential problems involves a lot of planning and work but it is critical for the success of the study. Appropriate members of the research team should remain in constant touch with the operational people for the duration of the study.

XIV. THE TENTH STEP: TESTING THE STUDY METHOD BEFORE TAKING FINAL DATA

In the past the process of testing the study method before taking the final data has been called "pretesting." That label tends to be slightly misleading, however, being confused with the process of early and preliminary testing of issues that are to be the basis for the transfer study. In any event, the process should consist of what amounts to a small dress rehearsal conducted before the actual study begins, the effort being an attempt to discover method problems that had not been predicted earlier.

As in other types of research, testing the study method is essential. It is indeed rare that all problems are predicted, regardless of the amount of care that has been devoted to the plan. Such method testing should be conducted sufficiently early to provide the research team with adequate time to make last minute fixes or corrections. Frequently the method testing process need use only a very few students who go through the entire course of the planned study. Possibly greater emphasis should be placed on routines involving experimental students; although routines for control students must not be ignored.

A problem may involve availability of students in sufficient numbers to conduct both the method testing work and the actual study. Depending on the number and severity of method problems discovered

(with changes being required for routines of the actual study), it is generally a good idea to provide that performance data from students used in method testing are not included with data from students in the factual study. Thus, the problem is one of not using too many of the limited number of students who are of a slightly different nature than those to be used in the actual study, although truly severe differences could pose a real problem. As is the case with many other issues for these transfer studies, the research team will have to exercise considerable imagination and judgment when and if the student scarcity problem is encountered.

XV. THE ELEVENTH STEP: ANALYSIS OF RESULTS

While the details of the data analyses will depend on the nature of the specific transfer study, a few observations can be made that should apply to many types of studies. As has been suggested, transfer studies should be concerned with reasonably substantial performance differences between or among groups of experimental and control students—differences that could have practical meaning. Interpretation of findings of a study should not be based solely on probability (p) levels associated with inferential tests for statistical significance because those p levels simply do not tell the entire story.

It is recommended that the first step of the analysis involve placing the raw performance data in one or more display formats that facilitate inspection. Inspection of those data should be made before, during, and subsequent to running inferential tests of interest. Such an inspection can perform several valuable services. First, if large performance differences exist, they will be evident by simply looking at the data. An inspection should be directed toward looking for both large group performance mean differences and variation of performances within the various groups. If performance variationis quite large, the use of arithmetic means to describe group performances is not entirely satisfactory without additional descriptors. For example, a large standard deviation for an array of values indicates that the array mean should not be taken too seriously. The wide variation of the individual values likely has considerable meaning that should be explored. Second, inspection of the raw data display formats during and after running statistical inferential tests will permit an understanding of the results of those tests.

As the data are analyzed using inferential tests, the results of those tests—as in an analysis-of-variance summary table—should be cross-compared with the raw data display formats, again with the understanding that probability levels do not tell the entire story. In conjunction with an analysis of variance summary, for example, it is highly useful to derive estimates of strengths of associations, such as simple values of eta squared or estimated omega squared. (For a discussion of the estimated omega squared statistic, see Hays, 1973, pp 484-488, 512-513). Perhaps the easiest way to see how these statistics are of value involves the descriptive eta squared (estimated omega squared being its inferential counterpart). Simply divide each of the sums of squares for main effects, interactions, and error by the total sum of squares, arriving at estimates of proportions of total variation that are accounted for by each. If eta squared for error is large, attention is directed to the variation of individual students' scores within arrays of the display of raw values, where it will be seen that there is not a great deal of uniformity of performances within those arrays. This finding would indicate that any statistically significant transfer effect should not be taken too seriously; i.e., the differences among student performances are more marked than differences among group means.

On the other hand, if the greater proportion of variation is associated with, say, main effects or interaction effects, i.e., the values of eta squared are relatively large, an inspection of the raw data will show that performance within arrays is reasonably uniform and that mean-differences among groups, which are of principal interest, represent strong effects. In other words, the larger the estimate of strength of association for main effects or interaction effects, the more credible are the results—p levels notwithstanding.

While it is unfortunate that many available computer programs do not provide for calculation of these values of strength of association, it is a relatively easy matter to calculate them "by hand" or to

provide that simple subroutines be added to those programs to present this critically important information.

In ending this discussion, it should be noted that, within limits, undue concern with underlying assumptions of parametric tests is incorrect, as is the insistence that parametrics be used only with data associated with interval or ratio scales. These fallacies take away the researcher's most powerful and versatile inferential tools. The notion of "robustness" of parametrics in terms of departures from assumptions of normality and homoscedasticity, careful interpretations of the assumption of data independence, and scales of measurement is discussed by Baker, Hardyck, & Petrinovick, 1970; Boneau, 1960, 1961; Burke, 1953; Hays, 1973; and Lord, 1953. The excessive use of nonparametric tests also is to be avoided because these tests tend to throw away large portions of the data and, in general, are characterized by relatively low power (e.g., they might not reject a false null hypothesis).

XVI. SOME CLOSING REMARKS

The goal of studies of transfer of learning is to provide information about techniques or equipment; the use of which can serve as guides for designing or updating training curricula. The likelihood that the information will be used depends on the extent to which both study method and results are convincing to the personnel responsible for operational training. Studies demonstrating large performance effects resulting from simulator pretraining certainly will be the most convincing and, other things being equal, will be the most likely to result in the use of experimental techniques or equipment during operational training.

This report has discussed a number of issues concerned with research methods, with emphasis on the need for careful planning. It has addressed definitions of the problem and the task, considerations of students, instructors, performance measurement, time requirements, dilutant factors, scheduling, the busy operational environment, method testing, and analysis of results. These issues provide the means by which the researcher can attempt to conduct a study illustrating the maximum possible transfer estimate for the task at hand, illustrating for the operational instructor what can be accomplished.

It is hoped that the researcher, viewing all of these issues in the aggregate, will not arrive at the unfortunate conclusion that it is virtually impossible to run a truly effective study of transfer of learning. Certainly no single study is likely to be able to observe all of the issues in their absolute form. But to the extent that a great many issues are taken into account, to that same extent the transfer study is likely to provide sound and useful results of benefit to the operational training community.

REFERENCES

Baker, B.O., Hardyck, C.D., & Petrinovich, L.F. Weak measurements vs strong statistics: An empirical critique of S. S. Stevens' proscriptions on statistics. *Educational and Psychological Measurement*, 1966, **26**(2), 291—309. (In: Haber, A., R. P. Runyon, and P. Badia (Eds.). *Readings in statistics*. Reading, MA: Addison-Wesley Publishing Co., 1970.)

Boneau, C.A. A note on measurement scales and statistical tests. *American Psychologist*, 1960, **16**, 260—261. (In: Haber, A., *et al.*, op. cit.)

Boneau, C.A. The effects of violations of assumptions underlying the *t* test. *Psychological Bulletin*, 1960, 57, 49—64. (In: Haver, A., *et al.*, op. cit.)

Burke, C.J. Additive scales and statistics. *Psychological Review*, 1953, **60**, 73—75. (In: Haber, A., *et al.*, op. cit.)

Eliis, H.C. *The transfer of learning*. New York: The Macmillan Company, 1965, 11—14.

Gagne, R.M., Forster, H., & Crowley, M.E. The measurement of transfer of training. *Psychological Bulletin*, 1948, **45**, 97—130.

Hays, W.L. *Statistics for the social sciences* (2nd Ed.). New York: Holt, Rinehart and Winston, Inc., 1973, 419, 429, 481—482, 484—488, 512—513.

Lord, F.M. On the statistical treatment of football numbers. *American Psychologist*, 1953, **8**, 750—751. (In: Haber, A., *et al.*, op. cit.)

McNemar, Q. Sampling in psychological research. *Psychological Bulletin*, 1940, **37**, 340.

Murdock, B.B., Jr. Transfer designs and formulas. *Psychological Bulletin*, 1957, **54**, 313—326.

Northrop Corporation, Aircraft Division. *Experiments to evaluate advanced flight simulation in air combat pilot training. Vol. 1: Transfer of learning experiment.* Contract No. N62269-74-C-0314. Final Report. Warminster, Pennsylvania: Naval Air Development Center, March 1976.

Payne, T.A., Dougherty, D.J., Hasler, S.G., Skeen, J.R., Brown, E.L., & Williams, A.C., Jr. *Improving landing performance using a contact landing trainer.* Technical Report SPECDEVCEN 71-16-11. Port Washington, New York: U.S. Navy, Special Devices Center, March 1954.

Roscoe, S.N. Incremental transfer effectiveness. *Human Factors*, 1971, **13**(6), 561—567.

Roscoe, S.N. A little more on incremental transfer effectiveness. *Human Factors*, 1972, **14**(4), 363-364.

Villiams, A.C., Jr., & Flexman, R.E. *An evaluationof the SNJ operational trainer as an aid in contact flight training.* Technical Report 71-16-5. Port Washington, New York: Office of Naval Research, Special Devices Center, July 1949.